Trend Assessment in a Long Memory Dependence Model using the Discrete Wavelet Transform

Peter F. Craigmile[1], **Peter Guttorp**[2] **and Donald B. Percival**[3].

[1] Department of Statistics, 404 Cockins Hall, 1958 Neil Avenue, The Ohio State University, Columbus. OH 43210.

[2] Department of Statistics, Box 354322, University of Washington, Seattle. WA 98195–4322.

[3] Applied Physics Laboratory, Box 355640, University of Washington, Seattle. WA 98195–5640.

Abstract

In this paper we consider trend to be smooth deterministic changes over long scales, and tackle the problem of trend estimation in the presence of long memory errors (slowly decaying autocorrelations). Using the fractionally differenced (FD) process as a motivating example of such a long memory process, we demonstrate how the discrete wavelet transform (DWT) is a natural choice at extracting a polynomial trends from such an error process. We investigate the statistical properties of the trend estimate obtained from the DWT, and provide pointwise and simultaneous confidence intervals for the estimate. Based on evaluating the power in the trend estimate relative to the estimated errors, we provide a test of nonzero trend. We finish by applying the methods to a climatological example.

Keywords: trend assessment; long memory dependence; fractionally differenced process; discrete wavelet transform; time series analysis.

1 Introduction

Trend assessment is an important problem in time series analysis. Even though trend is discussed in most introductory treatments on time series analysis, the concept of trend can be confusing. There is no commonly accepted precise definition for trend, but, to quote from Kendall (1973), "*the essential idea of trend is that it shall be smooth.*" In keeping with this viewpoint, we take trend to

be continuous slowly varying changes in a time series over long scales. These changes are presumed to be due to nonstochastic mechanisms, and hence trend is usually modeled independently of the stochastic portion of the series. These restrictions on the form that trend can take distinguish trend assessment from the more general problem of signal estimation (see, e.g., Johnstone and Silverman (1997) and Johnstone (1999) for details on the latter case).

In this paper we consider a time series Y_t that can be modeled as $Y_t = T_t + X_t$, where T_t is a nonstochastic trend component, while X_t is stochastic. Trend assessment is the problem of determining whether or not T_t is actually present in the time series; i.e., we wish to assess the null hypothesis that $T_t = 0$. We will make this assessment under the assumption that the stochastic component X_t is either a stationary long memory process or a nonstationary process whose backward differences of a certain order form a stationary time series. These forms for X_t make trend assessment particularly challenging because the stochastic component will have significant low frequency components that will be difficult to distinguish from a smoothly varying trend. The assessment of trend within the context of stochastic models that can produce trend-like realizations has been advocated by Smith (1993), p.143, who notes that what is really of interest is "*the separation of effects due to natural variability, such as may for example be represented by a stationary time-series model, and those that are due to a trend in the data*" (this quote is given in reference to the investigation of global warming, but is clearly applicable to other areas). This approach to trend assessment is conservative in that we will not falsely declare there to be a significant trend in the data if in fact the stochastic element is reasonably capable of generating the observed low frequency variations. This notion of trend and its assessment clearly has application in areas such as atmospheric science, hydrology, climatology and economics.

In what follows, we make the assumptions that the trend T_t is well approximated at least locally by a low order polynomial (such as a linear or quadratic) and that the stochastic component X_t is a fractionally differenced (FD) process. Under these assumptions, the discrete wavelet transform (DWT) that is based upon the Daubechies family of wavelet filters can be used to transform the time series Y_t into components that are attributable mainly to either T_t or X_t. The ability of the DWT to cleanly separate Y_t into these components allows us to propose (a) a simple estimator for T_t, (b) a test for its significance and (c) confidence bands for the unknown trend.

Various aspects of the approach we take to trend estimation have been addressed previously in the literature. Nicholls, Heathcote, and Cunningham (1986) investigate the use of smoothers and least square estimators to evaluate trends, but do not consider the stochastic component of the time series in detail. Künsch (1986) demonstrates it is possible using the periodogram to asymptotically discriminate between a weakly dependent process with a small monotonic trend and a stationary long memory process. Sibbertsen (2003) shows via Monte Carlo methods that it is possible to make this discrimination for finite sample sizes using a tapered log periodogram estimate. In the area of environmental statistics, Smith (1989b, 1989a) employs extreme value theory to analyze trend in ground level ozone series. In subsequent work Smith (1993) adapts the standard least squares estimator to test for trend (global warming) in temperature time series (we use a similar least squares estimator as a basis for power comparisons later in this paper). Brillinger (1994, 1996) uses the continuous Haar wavelet transform to provide pointwise confidence intervals for a trend estimate. We conjecture that his approach can be extended to other wavelet transforms if the associated wavelet function has bounded variation. Teverovsky and Taqqu (1997) consider tests for long memory dependence in the presence of two types of trend (shifting means and slowly decaying trend). By investigating the theory of cumulative sums the authors illustrate via Monte Carlo studies that the estimation bias is larger in the presence of trend, but that the standard deviations obtained are comparable. The authors provide no test for trend or confidence intervals. Deo and Hurvich (1998) investigate statistical inference for a linear trend in the presence of fractionally integrated errors with long memory parameter, $d \in [0, 1.5)$. Under a moment condition on the innovations of the process (which is stricter for $d \in (0.5, 1.5)$), they show that the ordinary least squares (OLS) estimate of the slope parameter is asymptotically normal for $d \neq 0.5$. This extends a result due to Yajima (1988, 1991) for $d \in [0, 0.5)$. The authors also consider two other estimators of the trend: the sample mean of the first order differences and the tapered sample mean of the first order differences. Asymptotic normality of these estimators is proved, and the three estimators of trend are compared by looking at the asymptotic variances and relative efficiencies. Beran (1999) investigates trend in the more general framework of an autoregressive fractionally integrated time series model. The trend is removed via a variable bandwidth smoother and then standard likelihood methods are used to estimate the parameters. The author iterates to handle the choice of bandwidth. Gilbert (1999) also tests for the onset of trend using the DWT but restricts analysis to white noise or AR(1) error sequences.

Previously (Craigmile, Percival, and Guttorp (2000b, 2000a)) we considered parameter estimation of polynomial trend contaminated FD processes. By assuming simplifying assumptions for the wavelet covariance structure we were able to estimate the parameters of the error process semiparametrically using a standard Gaussian likelihood model. This method has some advantages, namely flexibility to the underlying model. In this paper we consider assessment of trend using the DWT. In section 2 we provide a background on the DWT, and in section 3 we use the DWT todecompose the data into a trend and a stochastic component. In section 4 we investigate the statistical properties of the trend estimate. We provide a test for trend in section 5 and compare our test to that of a standard test of nonzero slope in a linear regression model. We consider a climatological example in section 6, and finish with a discussion in section 7.

2 Background on the discrete wavelet transform

In this section we define the discrete wavelet transform and state some of its key properties. For details, see, e.g., Chapter 4 of Percival and Walden (2000).

For an even positive integer L, let $\{h_l : l = 0, \ldots, L-1\}$ denote the Daubechies *wavelet filter* of unit l_2 norm and $\{g_l = (-1)^{l+1} h_{(L-1)-l} : l = 0, \ldots, L-1\}$ be the associated *scaling filter*. The squared gain functions (i.e., the squared modulus of the transfer function for each filter) for the wavelet and scaling filters are given by

$$\mathcal{H}_{1,L}(f) \;=\; \left| \sum_{l=0}^{L-1} h_l \, e^{-i2\pi f l} \right|^2 \;=\; 2 \, \sin^L(\pi f) \sum_{l=0}^{L/2-1} \binom{L/2-1+l}{l} \cos^{2l}(\pi f), \tag{1}$$

and $\mathcal{G}_{1,L}(f) = \mathcal{H}_{1,L}(1/2 - f)$, respectively. For a particular choice of L there are multiple filters $\{h_l\}$ and $\{g_l\}$ that share these squared gain functions. Daubechies (1992) distinguishes between two (of the possible) choices.

- The extremal phase, D(L), filters are the ones that exhibit the smallest delay (have maximum cumulative energy) over other choices of the scaling filter.
- The least asymmetric, LA(L), filters (defined for L=8, 10, ...) are the ones closest to a linear phase filter.

The discrete wavelet transform (DWT) can be defined efficiently in terms of a pyramid algorithm involving the wavelet and scaling filters (Mallat (1989)). Let N be a positive integer that is an integer multiple by 2^J for some positive integer J. Let $V_{0,t} = Y_t$ $(t = 0, \ldots, N-1)$ and $N_j = N2^{-j}$ $(j = 1, \ldots, J)$. Then the scaling coefficients for level j are defined recursively by

$$V_{j,k} = \sum_{l=0}^{L-1} g_l \, V_{j-1,2k+1-l \bmod N_{j-1}} \qquad k = 0, \ldots, N_j - 1.$$

These coefficients are associated with averages on scale 2^j and with times spaced 2^j units apart. Similarly the wavelet coefficients for level j are defined by

$$W_{j,k} = \sum_{l=0}^{L-1} h_l \, V_{j-1,2k+1-l \bmod N_{j-1}} \qquad k = 0, \ldots, N_j - 1.$$

These coefficients are associated with changes in averages on scale 2^{j-1} and with times spaced 2^j apart. The first $B_j = \lceil (L-2)(1-2^{-j}) \rceil$ wavelet coefficients are affected by circularly filtering; i.e., they incorporate data from the start and end of the sequence. We refer to these as the boundary coefficients. The remaining $M_j = N_j - B_j$ are called the nonboundary coefficients.

We can calculate the wavelet and scaling coefficients directly from Y_t using (page 152 of Percival and Walden (2000))

$$V_{j,k} = \sum_{l=0}^{L_j-1} g_{j,l} Y_{2^j(k+1)-1-l \bmod N}, \qquad W_{j,k} = \sum_{l=0}^{L_j-1} h_{j,l} Y_{2^j(k+1)-1-l \bmod N}, \tag{2}$$

for $k = 0, \ldots, N_j - 1$, where $\{g_{j,l}\}$ and $\{h_{j,l}\}$ denote the equivalent level j scaling and wavelet filters respectively of length $L_j = (2^j - 1)(L-1) + 1$. For example the level j Haar scaling filter is (page 103 of Percival and Walden (2000))

$$g_{j,l} = \begin{cases} 1/2^{j/2}, & l = 0, \ldots, 2^j - 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The squared gain functions for the level j scaling and wavelet filters are

$$\mathcal{G}_{j,L}(f) = \prod_{k=0}^{j-1} \mathcal{G}_{1,L}(2^k f) \quad \text{and} \quad \mathcal{H}_{j,L}(f) = \mathcal{H}_{1,L}(2^{j-1} f) \prod_{k=0}^{j-2} \mathcal{G}_{1,L}(2^k f).$$

It can be shown (chapter 4 of Percival and Walden (2000)) that the level j scaling filter acts as an approximate bandpass filter with passband $[0, 2^{-(j+1)}]$, and that the level j wavelet filter acts as an approximate bandpass filter with passband $[2^{-(j+1)}, 2^{-j}]$.

Formally the DWT of $\{Y_t\}$ is defined via a matrix multiplication that yields a vector whose elements are given by $\{W_{j,k}\}$ and $\{V_{J,k}\}$. Let

$$\mathbf{W} \;=\; (W_{1,0},\ldots,W_{1,N_1-1},W_{2,0},\ldots,W_{2,N_2-1},\ldots,W_{J,0},\ldots,W_{J,N_J-1},V_{J,0},\ldots,V_{J,N_J-1})^T,$$

and define $\mathcal{W}$ to be the $N \times N$ matrix such that $\mathbf{W} = \mathcal{W}\mathbf{Y}$ (see section 4.6 of Percival and Walden (2000) for the form of $\mathcal{W}$). For the Daubechies wavelet filters used here, $\mathcal{W}$ is orthogonal which implies that the inverse DWT is given by $\mathbf{Y} = \mathcal{W}^T\mathbf{W}$. We can also define the inverse transform using a pyramid algorithm. Given the wavelet and scaling coefficients for level j, the previous level scaling coefficients can be recovered using

$$V_{j-1,k} \;=\; \sum_{l=0}^{L-1}\left[g_l\, V^{(2)}_{j,k+l \bmod N_{j-1}} + h_l\, W^{(2)}_{j,k+l \bmod N_{j-1}}\right], \qquad k = 0,\ldots,N_j-1. \tag{4}$$

where $^{(2)}$ denotes the upsampling operator, e.g.,

$$W^{(2)}_{j,k} = \begin{cases} 0, & \text{if } k \text{ is even;} \\ W_{j,(k-1)/2}, & \text{otherwise.} \end{cases}$$

3 Using the discrete wavelet transform to decompose trend

Suppose we observe data, $\{Y_t : t = 0,\ldots,N-1\}$, following the model

$$Y_t \;=\; T_t + X_t, \tag{5}$$

where $\{T_t\}$ is a deterministic polynomial trend of order K, and $\{X_t\}$ is a realization of a zero mean Gaussian FD(d,σ^2) process with difference parameter d and innovation variance $\sigma^2 > 0$. By definition $\{X_t\}$ has a spectral density function, $S(f)$, given by

$$S(f) \;=\; \sigma^2|2\sin(\pi f)|^{-2d}, \qquad |f| \le 1/2. \tag{6}$$

For $d \in (-1/2,1/2)$ the process is stationary and invertible, and is a white noise (i.e., uncorrelated) process when $d = 0$. When $d \ge 1/2$ we obtain a class of non-stationary processes that become stationary after differencing $\lfloor d+1/2 \rfloor$ times. For $d \in [-1/2,1/2)$ the autocovariance sequence (acvs) can be shown to be (Granger and Joyeux (1980))

$$s_k(d,\sigma^2) \;=\; \sigma^2\,\frac{(-1)^k\,\Gamma(1-2d)}{\Gamma(1-d+k)\,\Gamma(1-d-k)}. \tag{7}$$

Beran (1994) lists further properties of FD processes.

Given an estimate of the FD process parameters (d and σ^2) we are interested in providing a wavelet estimate of $\{T_t\}$, and then testing the significance of the trend. A Daubechies wavelet filter of order $L \geq 2\max\{K+1, \lfloor d+1/2 \rfloor\}$ can decompose $\{Y_t\}$ into its constituent parts. The nonboundary coefficients contain a contribution due to $\{X_t\}$ alone, while the boundary wavelet coefficients and all the scaling coefficients contain all of $\{T_t\}$ (for details see Craigmile, Percival, and Guttorp (2000b)). In light of this it seems reasonable to use the inverse wavelet transform of these latter coefficients as an estimate of trend. We can view this as a "kill or keep" wavelet thresholding strategy. More precisely, let diag($\mathbf{x}$) denote the square matrix with $\mathbf{x}$ along the main diagonal and zeroes on the off-diagonals. For an integer n let $\mathbf{0}_n$ denote a vector of n zeroes, and $\mathbf{1}_n$ a vector of n ones. Let $A = \text{diag}(\mathbf{1}_{B_1}, \mathbf{0}_{M_1}, \mathbf{1}_{B_2}, \mathbf{0}_{M_2}, \ldots, \mathbf{1}_{B_J}, \mathbf{0}_{M_J}, \mathbf{1}_{N_J})$. Also let I_N be the $N \times N$ identity matrix. Since the DWT is an orthonormal transform, we can partition the data vector as

$$\mathbf{Y} \;=\; \mathcal{W}^T\mathbf{W} \;=\; \mathcal{W}^T A\mathbf{W} + \mathcal{W}^T(I_N - A)\mathbf{W} \;=\; \widehat{\mathbf{T}} + \widehat{\mathbf{X}}. \tag{8}$$

$\widehat{\mathbf{T}}$ captures the continuous changes in longer scales (the *trend*), in addition to long range stochastic variability. Since we are adding in the contribution due to the boundary wavelet coefficients, the estimate also contains some short scale variability at the start and end points and can thus be said to have a variable bandwidth; on the other hand, $\widehat{\mathbf{X}}$ will be a bandpass version of $\mathbf{X}$ with an approximate bandpass frequency of $[2^{-(J+1)}, 2^{-1}]$ (a combination of the bandpasses $[2^{-(j+1)}, 2^{-j}]$ for each wavelet level $j = 1, \ldots, J$.) Let us illustrate the properties of $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{T}}$ using the following example. Let

$$Y_t \;=\; (20(t/512 - 0.5)^2 - 2) + X_t, \qquad t = 0, \ldots, 511,$$

where $\{X_t\}$ is a realization of an FD(0.4,1) process. A realization of $\{Y_t\}$ is shown in panel (a) of Figure 1, with the true trend being the solid line. We decompose the data using a LA(8) wavelet filter analyzing to level $J = 7$. The estimate of the trend, $\widehat{\mathbf{T}}$, is the jagged line shown in panel (b), with the true trend again being the solid line. The estimate of trend oscillates around the true value with more variability at the ends of the series. Panels (c) and (d) show the estimated residuals, $\widehat{\mathbf{X}}$ and the actual realization of the error process $\mathbf{X}$. It is hard to pick up differences by comparing these two plots, and so the periodograms for the estimated and actual errors are shown in panels (e) and (f). The six dotted lines on each plot indicate the approximate bandpass frequencies $[2^{-(j+1)}, 2^{-j}]$

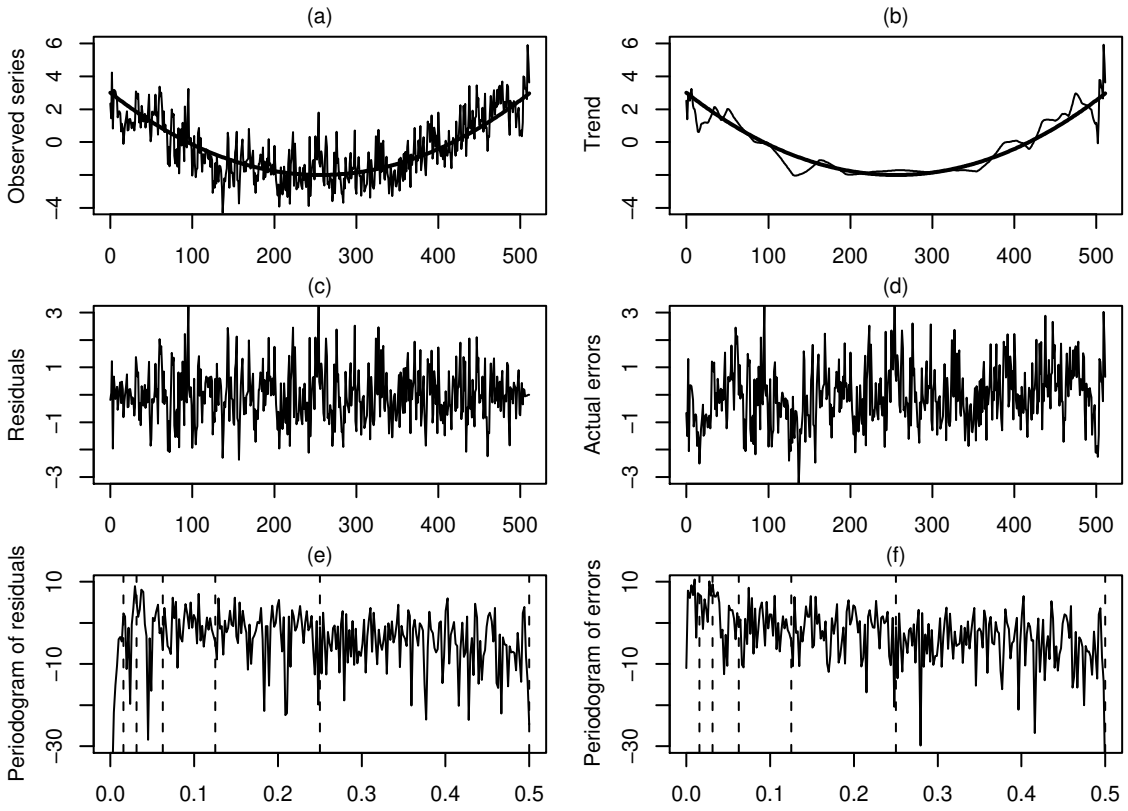

Figure 1: (a) Simulation of a quadratic trend plus an FD(0.4,1) process ($N = 512$) (jagged line) and the true trend (solid line); (b) The estimated (jagged line) and true trend (smooth line); (c) The estimated residuals; (d) The actual errors; (e) Periodogram of the estimated residuals. The vertical dotted lines indicate the approximate bandpass frequencies, $[2^{-(j+1)}, 2^j]$ for each wavelet level $j = 1, \ldots, 5$; (f) Periodogram of the actual realization of the error process.

for the wavelet filters on level $j = 1, \ldots, 5$. We can see that the spectral content of the estimated and actual errors is virtually identical at the higher frequencies.

4 Statistical properties of the trend estimate

Let $\text{cov}(\mathbf{X})$ denote the $N \times N$ covariance matrix for the vector $\mathbf{X}$, which is multivariate Gaussian with a mean vector of $\mathbf{0}_N$. Since

$$\widehat{\mathbf{T}} = \mathcal{W}^T A \mathbf{W} = \mathcal{W}^T A \mathcal{W} \mathbf{Y} = R\mathbf{T} + R\mathbf{X} \text{ with } R = \mathcal{W}^T A \mathcal{W},$$

standard theory says that $\widehat{\mathbf{T}}$ is multivariate Gaussian with mean $R\mathbf{T}$ and covariance $\mathrm{cov}(\widehat{\mathbf{T}}) = R\,\mathrm{cov}(\mathbf{X})R^T$. We consider calculating these covariance matrices in some detail. For $d \in [-1/2, 1/2)$, $\{X_t\}$ is a stationary FD process, and $\mathrm{cov}(\mathbf{X})$ is a symmetric Toeplitz matrix, simplifying the calculations somewhat. For $d \geq 1/2$ the FD process is not stationary, but since the process is stationary after differencing $\lfloor d + 1/2 \rfloor$ times we can still calculate the elements of the covariance matrix as an aggregation of the covariances in the $d \in [-1/2, 1/2)$ case. As an example consider an FD process with $d \in [1/2, 3/2)$. The first order difference of this process is a stationary process and

$$\mathrm{cov}(X_s, X_t) \;=\; \sum_{j=1}^{s} \sum_{k=1}^{t} s_{j-k}(d-1, \sigma^2), \qquad s, t = 0, \ldots, N-1,$$

using the acvs given by equation (7). The variance will be large as d increases above a half.

We now investigate the bias and variance properties of the trend estimator. The first proposition deals with the bias in estimating $\mathbf{T}$ by $\widehat{\mathbf{T}}$.

Proposition 4.1 *Suppose that $L \geq 2(K+1)$. Then under model (5) the bias in estimating $\boldsymbol{T}$ by $\widehat{\boldsymbol{T}}$ is zero.*

Proof Since a Daubechies (1992) wavelet filter has $L/2$ embedded differencing operations (pages 105–106 of Percival and Walden (2000)), the nonboundary wavelet coefficients of a zero mean FD process plus polynomial trend have zero mean. Thus a linear combination of the nonboundary wavelet coefficients such as $\widehat{\mathbf{X}}$ also has zero mean. Applying this to model (5),

$$E(\widehat{\mathbf{T}} - \mathbf{T}) \;=\; E((\mathbf{Y} - \widehat{\mathbf{X}}) - (\mathbf{Y} - \mathbf{X})) \;=\; E(\mathbf{X} - \widehat{\mathbf{X}}) \;=\; \mathbf{0}_N,$$

since $E(\mathbf{X}) = \mathbf{0}_N$.

Let $\widehat{T}_t$ be the tth element of the vector $\widehat{\mathbf{T}}$. We shall now derive an expression for $\mathrm{var}(\widehat{T}_t)$. From the definition of $\mathrm{cov}(\widehat{\mathbf{T}})$ we have

$$\mathrm{var}(\widehat{T}_t) \;=\; \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R_{t,i}\, \mathrm{cov}(X_i, X_j)\, R_{t,j},$$

where $R_{t,i}$ is the (t,i) element of the matrix $R = \mathcal{W}^T A \mathcal{W}$; i.e.,

$$R_{t,i} \;=\; \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \mathcal{W}_{m,t}\, A_{m,n}\, \mathcal{W}_{n,i} \;=\; \sum_{m=0}^{N-1} A_{m,m}\, \mathcal{W}_{m,t}\, \mathcal{W}_{m,i},$$

since A is a diagonal matrix (see section 4.6 of Percival and Walden (2000) for details on calculating $\mathcal{W}$ based on the wavelet filter, h_l, and the scaling filter, g_l). When the FD process is stationary we can calculate the variance from the following equation

$$\text{var}(\widehat{T}_t) \quad = \quad \sum_{k=-(N-1)}^{N-1} s_k(d,\sigma^2)\, r_{t,k},$$

where we define $r_{t,k} = \sum_{i=0}^{N-1-|k|} R_{t,i} R_{t,i+|k|}$, and we calculate the acvs recursively using

$$s_k(d,\sigma^2) \quad = \quad \begin{cases} \sigma^2 \, \dfrac{\Gamma(1-2d)}{\Gamma^2(1-d)}, & k=0; \\ \\ s_{k-1}(d,\sigma^2) \left[\dfrac{k-1+d}{k-d} \right], & k \geq 1. \end{cases}$$

Overall, this is an $O(N^3)$ calculation (given we have already calculated $\mathcal{W}$), but can be calculated as an $O(N^2 \log N)$ calculation by using DFTs to calculate $r_{t,k}$ (for an example of how this can be done see exercise 3.14 on page 123 of Percival and Walden (1993)).

The variance is proportional to σ^2 and depends on d through a ratio of gamma functions. From the dependence on $\mathcal{W}$, the variance of the trend estimate is affected by the choice of wavelet filter and the choice of the number J of wavelet levels to analyze. The variance also depends on the sample size N. For the Haar wavelet filter, we can state the following explicit result.

Proposition 4.2 *Suppose we analyze to $J \leq \log_2(N)$ levels using the Haar wavelet filter. Then the variable bandwidth trend estimate at time $t = 2^J r + s$ for $r = 0, \ldots, N_J - 1$ and $s = 0, \ldots, 2^J - 1$ is given by*

$$\widehat{T}_t \quad = \quad 2^{-J} \sum_{l=0}^{2^J-1} Y_{2^J r + l}.$$

with a variance of

$$\text{var}(\widehat{T}_t) \quad = \quad 2^{-2J} \sum_{l=0}^{2^J-1} \sum_{l'=0}^{2^J-1} s_{l-l'}(d,\sigma^2),$$

that is constant for all t.

Proof With the Haar wavelet filter $B_j = 0$ for all $j = 1, \ldots, J$. Thus there are no boundary wavelet coefficients, and the trend estimate is the inverse DWT of the level J scaling coefficients.

Using equations (2) and (3) the level J scaling coefficients are given by

$$V_{J,k} = \frac{1}{2^{J/2}} \sum_{l=0}^{2^J-1} Y_{2^J k+2^J-1-l \bmod N} = \frac{1}{2^{J/2}} \sum_{l=0}^{2^J-1} Y_{2^J k+l},$$

if we reverse the order of summation. By page 100 of Percival and Walden (2000), the last N_J rows of $\mathcal{W}$ are circularly shifted versions of $\{g_{J,l}\}$ periodized to length N. In particular letting $t = 2^J r + s$ for $r = 0, \ldots, N_J - 1$ and $s = 0, \ldots, 2^J - 1$

$$\mathcal{W}_{N-1-N_J+k,t} = \begin{cases} 2^{-J/2}, & 2^J k \le t \le 2^J k + 2^J - 1, \\ 0 & \text{otherwise} \end{cases}$$
$$= \begin{cases} 2^{-J/2}, & k = r, \\ 0 & \text{otherwise.} \end{cases}$$

Thus

$$\widehat{T}_t = \widehat{T}_{2^J r+s} = \sum_{k=0}^{N_J-1} \mathcal{W}_{N-1-N_J+k,2^J r+s} V_{J,k} = 2^{-J/2} V_{J,r} = 2^{-J} \sum_{l=0}^{2^J-1} Y_{2^J r+l},$$

and hence

$$\text{var}(\widehat{T}_t) = 2^{-2J} \sum_{l=0}^{2^J-1} \sum_{l'=0}^{2^J-1} s_{l-l'}(d, \sigma^2),$$

which is invariant of t.

Using the Haar wavelet filter the trend estimate is a local average of the data. As we decrease J we increase the variance of the trend estimate. Figure 2 illustrates the variance of the trend estimate for various N, J and d (on the $\log_{10}$ scale). Keeping the other parameters constant we have that the variance increases at a very fast rate with d and decreases with increasing N and/or J.

The variance of the trend estimate is quite different for $L > 2$ for two reasons. First, the influence of the boundary wavelet coefficients induces higher variability at the endpoints of the time series (this is similar to the behavior of confidence bounds for simple linear regression, e.g. Draper and Smith (1998), p.82, Figure 3.1). Second, the overall variability oscillates with time, dependent on the shape of the scaling filter. As an example see Figure 3 in which we use a D(4) wavelet with $N = 256$ and $J = 4, 5, 6$ ($J = 6$ is the largest wavelet level for which there are still nonboundary wavelet coefficients). Each panel displays a different combination of d and J. We can see that as in the Haar case, increasing the value of d increases the overall variability, but also magnifies the

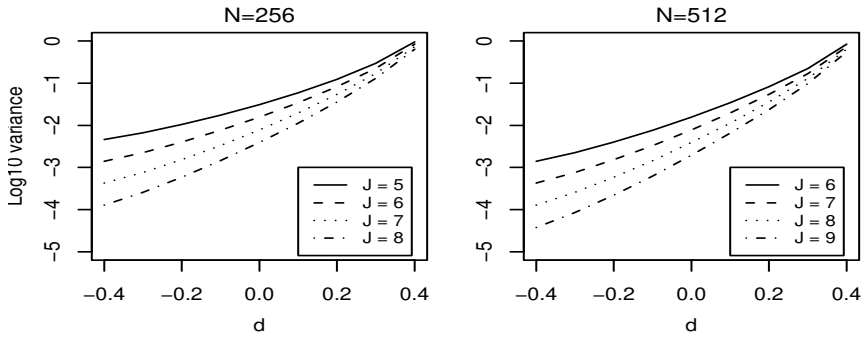

Figure 2: Plots of var($\widehat{T}_t$) (on the $\log_{10}$ scale) for the Haar wavelet filter at various values of d, J and N. The variance of the trend estimate is constant with time, t.

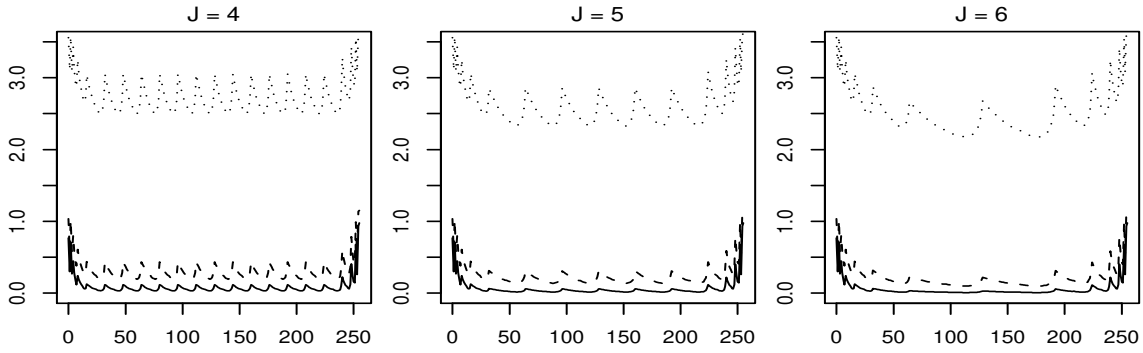

Figure 3: Plots of var($\widehat{T}_t$) by time index, t, for a realization of an FD(d,1) process of length 256 analyzed using a D(4) wavelet filter to level J. The solid curve is when $d = 0$, the broken curves is when $d = 0.25$, and the dotted curve is when $d = 0.45$.

amplitude of the oscillations. As we decrease J we decrease the wavelength of the oscillations for the central time points (those unaffected by the boundary wavelet coefficients) and the overall variance level for these points increases. We can also argue this latter result using the spectral decomposition theorem for stationary processes combined with the frequency localization result for wavelets. As we decrease J we increase the length of the passband of the signal which we capture in the trend estimate, and hence increase the variability of the trend estimate for the middle time points. The variances at the endpoints is relatively unchanged (for the same reason).

Finally we investigate how the variance depends on the choice of wavelet filter. Figure 4 plots the variances for the Haar, D(4), D(8) and LA(8) wavelet filters when $N = 256$ and $J = 5$, for various values of d. Comparing the filters we can see for the most part the general level of the variability increases with wavelet order (especially at the end points). We note that for fixed values of J there

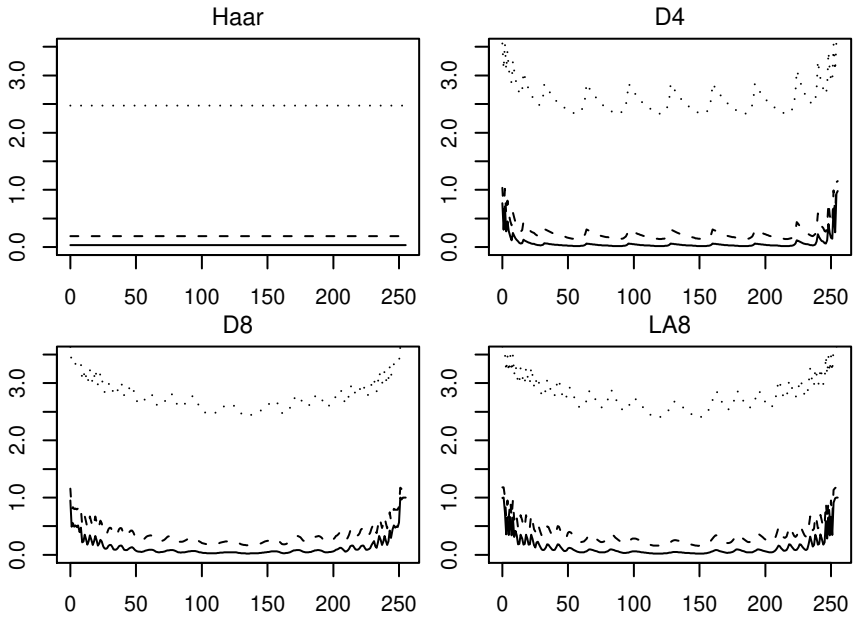

Figure 4: Plots of var($\widehat{T}_t$) by time index for the LA(8), D(8), D(4) and Haar wavelet filters, at various values of d, when $J = 5$ and $N = 256$. The solid curve is when $d = 0$, the broken curves is when $d = 0.25$, and the dotted curve is when $d = 0.45$.

are time indexes for which the variability is lower than the Haar, especially for the D(4) filter. Due to the shape of the wavelet filters the oscillations differ for the least asymmetric and the extremal phase.

In summary we have shown that the variable bandwidth estimate is an unbiased estimate of the true trend. When we analyze the data using a Haar wavelet ($L = 2$) the variance of the trend estimate is constant for all time points. For $L > 2$ the variance of the trend estimate changes with time. The variance oscillates depending on the value on N and J, and is larger at the beginning and ending time points. Regardless of the filter used, the variance of the trend estimate increases with increasing d and σ^2.

4.1 Pointwise confidence intervals

We now use the results of the previous section to derive pointwise confidence intervals (CIs) for the underlying trend. Let q_α denote the αth quantile of the standard Gaussian distribution. It follows

that a pointwise $100(1-\alpha)\%$ CI for the variable bandwidth trend estimate is given by

$$\widehat{T}_t \pm q_{1-\frac{\alpha}{2}} \sqrt{\mathrm{var}(\widehat{T}_t)}, \qquad t = 0 \ldots N-1. \tag{9}$$

In practice we estimate the FD process parameters and do not know the true value of $\mathrm{var}(\widehat{T}_t)$ for each t. We will examine this via Monte Carlo simulations. We first investigate the coverage probabilities in the case that the parameters of the FD process are considered known. Figure 5 shows a plot of the pointwise simulated coverage probabilities for the following combinations of parameters:

- Haar, D(4) and LA(8) wavelet filters;
- $d = 0, 0.25$ and 0.45;
- $N = 256$ and J determined by picking the largest J such that $M_J > 0$.

In each case we simulated an FD$(d, 1)$ process using the Davies–Harte algorithm (Davies and Harte (1987), Wood and Chan (1994), Craigmile (2002)), calculated the estimate of trend and associated 95% CI (assuming the known d and σ^2) and determined whether or not the true trend (a zero trend in this case) was contained in the interval for each time point. By repeating this procedure 2048 times we can calculate an estimate, $\hat{p}_t$, of the pointwise coverage probability for $t = 0, \ldots, N-1$. The solid curve denote these estimates, and the dotted 95% CIs are those for a standard binomial trial, i.e.,

$$\hat{p}_t \pm 1.96 \sqrt{\frac{\hat{p}_t(1-\hat{p}_t)}{N}}.$$

Note that the coverage probabilities are obviously correlated with time index. Although we miss the 95% mark occasionally, on the whole the figure verifies equation (9).

In practice we estimate the FD process parameters from the data. Suppose $\widehat{d}_N$ and $\widehat{\sigma}^2_N$ are estimates of the FD parameters. A CI for the trend estimate is then given by

$$\widehat{T}_t \quad \pm \quad q_{1-\frac{\alpha}{2}} \sqrt{\widehat{\mathrm{var}}(\widehat{T}_t)},$$

where $\widehat{\mathrm{var}}(\widehat{T}_t)$ denotes the estimated variance when we plug in the estimates of the FD process parameters into the expression for $\mathrm{var}(\widehat{T}_t)$. We now employ a Monte Carlo simulation to investigate the coverage probabilities for this situation (similar to known parameter case above). In our

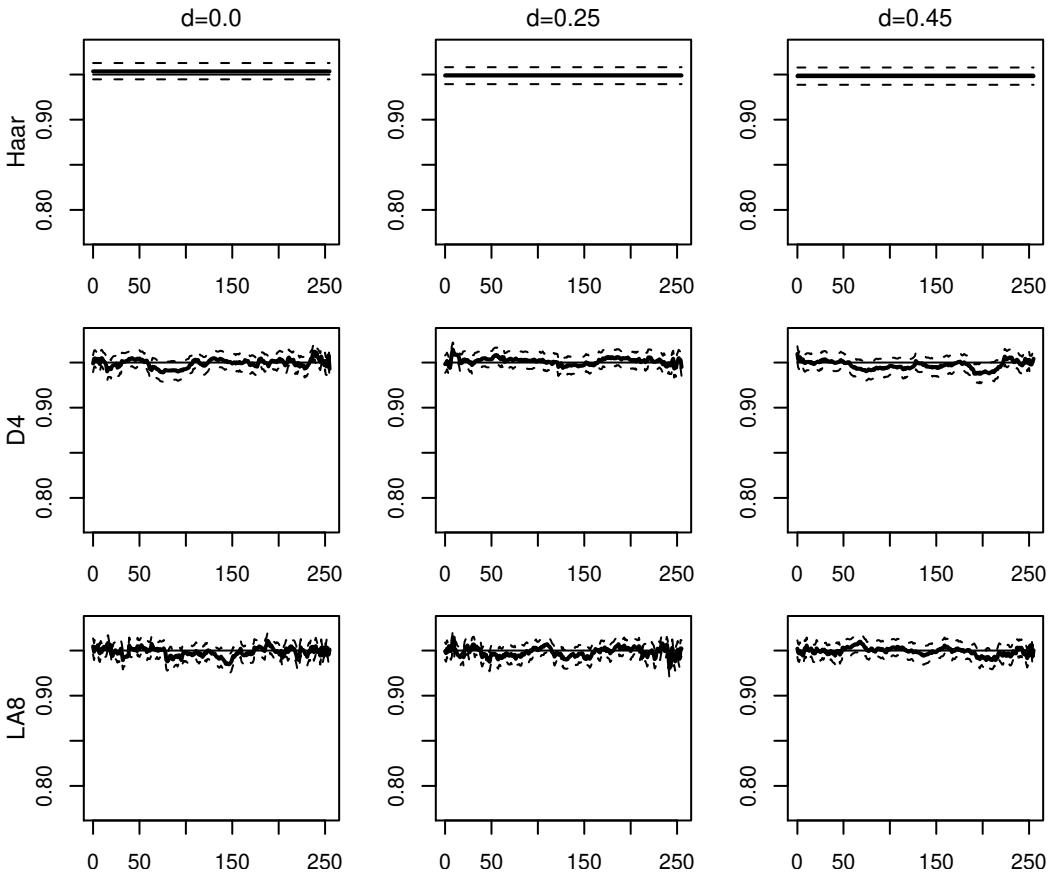

Figure 5: Simulated coverage probabilities for 95% pointwise CI for $\widehat{\mathbf{T}}$, for Haar, D(4) and LA(8) wavelet filters, $d = 0$, 0.25 and 0.45, $N = 256$, and J determined by picking the largest J such that $M_J > 0$. We assume a zero trend and calculate $\text{var}(\widehat{T}_t)$ based on the known values of d and $\sigma^2 = 1$. The calculations are based on 2048 Monte Carlo replications.

simulations we estimate d and σ^2 using the approximate white noise wavelet maximum likelihood model of Craigmile, Percival, and Guttorp (2000b), and then build a CI using the formula above. The results are shown in Figure 6. There is strong evidence that the coverage probabilities are not 95%. For the white noise case ($d = 0$) we obtain more conservative intervals, and for $d = 0.25$, 0.45 we have anti-conservative intervals. Indeed for d close to zero, it is more likely that we will have estimates of d which are negative. This leads to an antipersistent FD process (i.e., having negative autocorrelations at nonzero lags). This deflates the variance of the trend estimates, and increases the simulated coverage probabilities. As we increase the value of d, the average coverage probability decreases.

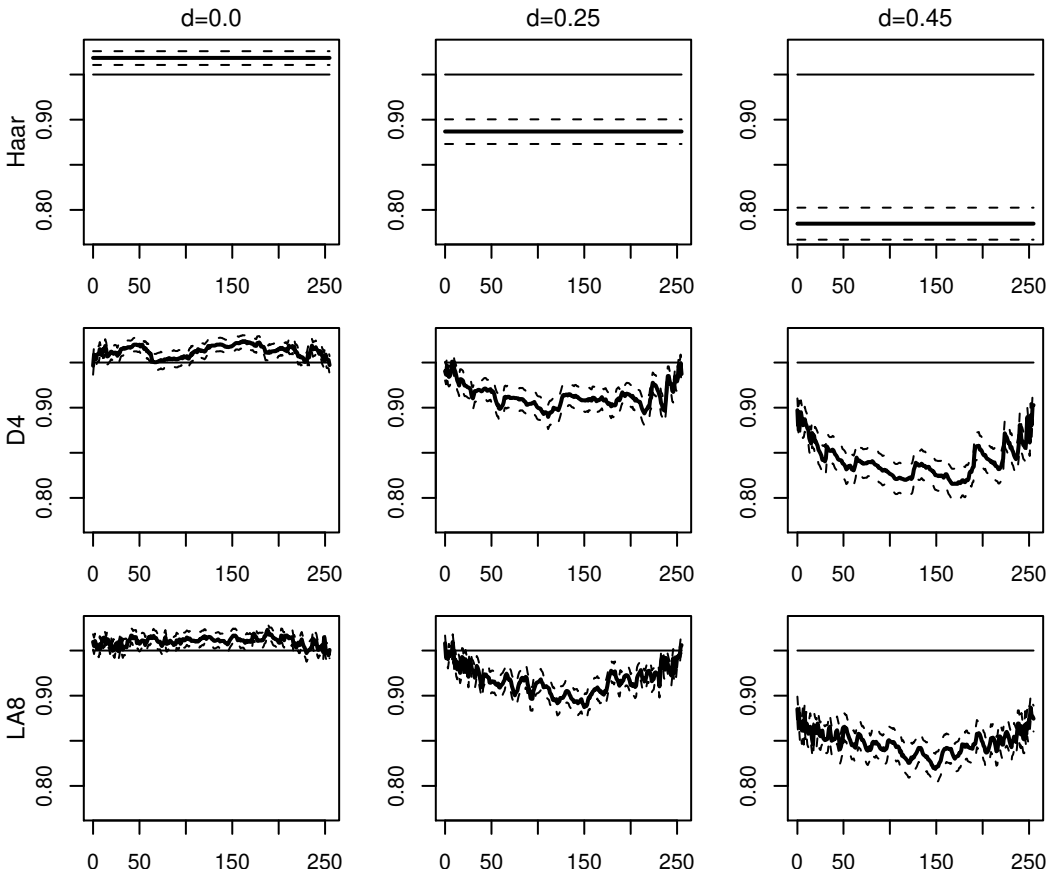

Figure 6: Simulated coverage probabilities for 95% pointwise CI for $\widehat{\mathbf{T}}$, for Haar, D(4) and LA(8) wavelet filters, $d = 0, 0.25$ and 0.45, $N = 256$, and J determined by picking the largest such J that $M_J > 0$. We assume a zero trend and calculate $\text{var}(\widehat{T}_t)$ based on estimating the value of d and σ^2. The calculations are based on 2048 Monte Carlo replications.

To improve the coverage probabilities we can consider schemes for providing a better estimate of $\widehat{\text{var}}(\widehat{T}_t)$. One solution is to average the estimated variance of the trend estimate over a limiting distribution for the FD process parameter estimates. One Monte Carlo approximation is to sample $\hat{d}_N$ and $\widehat{\sigma}^2_N$ from their limiting distributions. For each sampled value of the parameter estimates we calculate $\widehat{\text{var}}(\widehat{T}_t)$, and then average these estimated variances under repeated sampling. Figure 7 displays the simulated coverages for samples from the asymptotic distribution of the approximate white noise wavelet model estimates (Theorem 6.1 of Craigmile, Percival, and Guttorp (2000b)). Comparing with Figure 6, we note that the coverages are larger in this case. Indeed for $d = 0$ the coverages are more conservative.

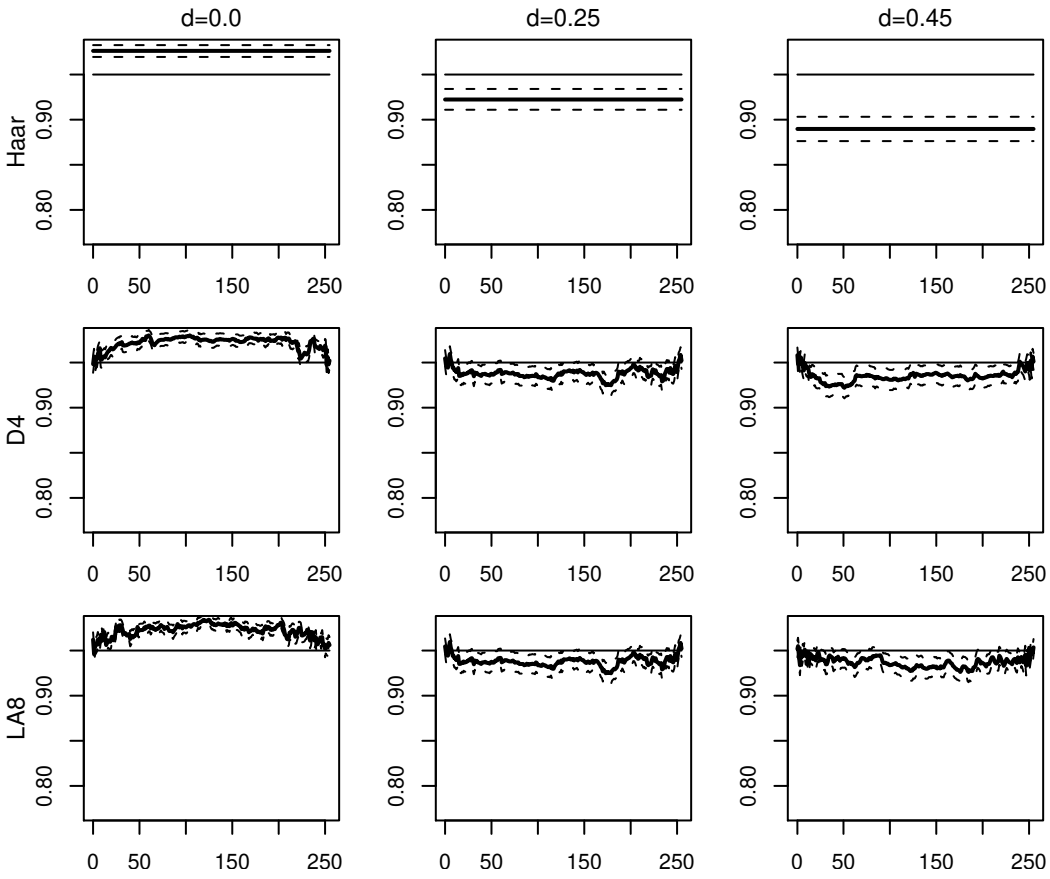

Figure 7: Simulated coverage probabilities for 95% pointwise CI for $\widehat{\mathbf{T}}$, for Haar, D(4) and LA(8) wavelet filters, $d = 0, 0.25$ and 0.45, $N = 256$, and J determined by picking the largest J such that $M_J > 0$. We assume a zero trend, estimate d and σ^2, and calculate expected variances based on the approximate limit theory for the parameter estimates. We used 2048 Monte Carlo replications.

In summary we have shown that it possible to construct pointwise CIs for the trend estimates. If we estimate the variance of the trend estimate using the FD process parameter estimates, the procedure compromises the coverage probabilities for the CI. By accounting for the variability in estimating the FD process parameters we can improve the coverage properties of the intervals.

4.2 Simultaneous confidence intervals

We now derive simultaneous $100(1-\alpha)\%$ CIs for $\mathbf{T}$. Let $\mathbf{v} = (\sqrt{\text{var}(\widehat{T}_1)}, \sqrt{\text{var}(\widehat{T}_2)}, \ldots, \sqrt{\text{var}(\widehat{T}_N)})^T$ and $\mathbf{U} = \mathbf{T} - \widehat{\mathbf{T}}$. We want to calculate the value of δ such that

$$1-\alpha \;=\; \Pr\Big(\widehat{\mathbf{T}} - \delta\mathbf{v} \le \mathbf{T} \le \widehat{\mathbf{T}} + \delta\mathbf{v}\Big) \;=\; 1 - 2\Pr(\mathbf{U} > \delta\mathbf{v}),$$

by the symmetry and continuity of the multivariate normal distribution. Hence we want δ such that $\alpha/2 \;=\; \Pr(\mathbf{U} > \delta\mathbf{v})$. Suppose we know the FD process parameters. Then $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, R\,\text{cov}(\mathbf{X})R^T)$, and we have $\alpha/2 = \Pr(\text{a } R\mathcal{N}(\mathbf{0}, \text{cov}(\mathbf{X})) \text{ r.v.} > \delta\mathbf{v}) = p(\delta)$, say. Clearly evaluation of δ in this case is not simple. We use a Monte Carlo approach. For some positive integer, I, and some fixed set of δ_j, $j = 1, \ldots, n_\delta$

- Set $n_j = 0$. For $i = 1, \ldots, I$:
  - Simulate an FD(d,σ^2) process and calculate $\widehat{\mathbf{T}}$;
  - If $\widehat{\mathbf{T}} > \delta_j$ $\mathbf{v}$ increment n_j by one.
- Let $\hat{p}(\delta_j) = n_j/I$.

Now choose the value of δ_j such that $\hat{p}(\delta_j)$ is closest to $\alpha/2$. We can use a divide and conquer algorithm since $p(\delta_j)$ increases with δ_j for fixed values of R, $\text{cov}(\mathbf{X})$ and hence $\mathbf{v}$.

5 A test for trend

In this section we explore a test for nonzero trend based on the DWT coefficients. Our hypotheses are

$$H_0 : T_t = 0 \text{ for all } t = 0, \ldots, N-1, \quad \text{versus} \quad H_1 : \text{not } H_0.$$

In keeping with the idea of assessing a trend in terms of the natural stochastic variability in a time series (Smith 1993), we propose a test for trend based upon a comparison of the variability in $\widehat{\mathbf{T}}$ and in $\widehat{\mathbf{X}}$. We define

$$P^*(d) \;=\; \frac{||\widehat{\mathbf{T}}||^2}{||\widehat{\mathbf{X}}||^2} \;=\; \frac{||A\mathbf{W}||^2}{||(I_N - A)\mathbf{W}||^2}, \tag{10}$$

by the analysis of variance relationship for the DWT coefficients (Percival and Walden (2000), section 4.6). Under H_0 we expect the variable bandwidth estimate to be close to zero, and thus we reject the null hypothesis for large values of $P^*(d)$. Using the analysis of variance relationship again

$$P^*(d) \quad = \quad \frac{||\mathbf{Y}||^2 - ||(I_N - A)\mathbf{W}||^2}{||(I_N - A)\mathbf{W}||^2} \quad = \quad \frac{||\mathbf{Y}||^2}{||(I_N - A)\mathbf{W}||^2} - 1.$$

Consequently we reject the null hypothesis for large values of the sum of squares of the trend relative to the sum of the squared nonboundary wavelet coefficients. Since the -1 term is unimportant to the outcome of the test, we let

$$P(d) \quad = \quad P^*(d) + 1 \quad = \quad \frac{||\mathbf{Y}||^2}{||(I_N - A)\mathbf{W}||^2}, \tag{11}$$

be our test statistic. An equally valid pair of hypotheses to consider is

$$H_0 : T_t = \text{some constant for all } t = 0, \ldots, N-1, \quad \text{versus } H_1 : \text{not } H_0.$$

In this case we do not test whether the trend is significant relative to zero, but relative to the mean of the signal. An obvious test statistic in this case is to center the data in the numerator yielding

$$P_c(d) \quad = \quad \frac{||\mathbf{Y} - \mathbf{1}_N \bar{Y}_N||^2}{||(I_N - A)\mathbf{W}||^2}.$$

This version of the statistic is often of more use in practice. While the sample mean is not as efficient an estimator of the process mean as the best linear unbiased estimator, it is asymptotically nearly as efficient as the latter for long memory processes and is much easier to compute (see section 8.2 of Beran (1994) for a comparison of different estimates of the process mean).

5.1 Distribution of the test statistic

The distribution of $P(d)$ (and $P_c(d)$) under the null hypothesis depends on the FD parameter, d, but is independent of σ^2, since both the numerator and denominator of the statistics have a factor of σ^2 which cancels out. Let χ^2_k denote a chi-squared random variable with k degrees of freedom. For Gaussian IID noise ($d = 0$), $\sum_{t=0}^{N-1} Y_t^2 \sim \sigma^2 \chi^2_N$, and since the DWT of a realization of a white noise process is itself a white noise process $\sum_{j=1}^{J} \sum_{k=B_j}^{N_j - 1} W_{j,k}^2 \sim \sigma^2 \chi^2_M$. Here the numerator and denominator are not independent random variables. The denominator depends on M which is defined in terms of N, J and L, so the distribution of $P(d)$ is dependent on these factors in addition

to d. For $d \in (0, 1/4]$, we can appeal to Beran (1994), Theorem 3.1 to obtain a Gaussian limit for the numerator of $P(d)$ (for large N), but clearly this is not general enough a result for all d. In practice a simple testing framework is to use a Monte Carlo test. We simulate a large number of FD(d,1) processes with no trend, and calculate the resulting test statistic. For any given time series the proportion of times the observed test statistic exceeds the simulated values is the P-value or observed level of significance.

Table 1 lists the upper quantiles of the distribution of $P_c(d)$ for various values of N, L and d. In each case we picked the maximum value of J such that there were a nonzero number of nonboundary wavelet coefficients on level J. The evaluation of the distribution was based on 10000 simulations of an FD process using the Davies–Harte algorithm and standard errors for a quantile at probability p, $F^{-1}(p)$ were calculated using the equation

$$\text{se}(F^{-1}(p)) \quad = \quad \sqrt{\frac{p\,(1-p)}{10000\,[f(p)]^2}}.$$

Here $f(p)$ denotes the density of the test statistic. We used the `density` command in S-PLUS to evaluate the density using a Gaussian window smoother. The value of $P_c(d)$ increases with d, and decreases with N. We see the same for the standard errors. As we increase L from 4 to 8, we increase the value of the test statistic. This is intuitive as we decrease the number of nonboundary wavelet coefficients that are included in the denominator sum.

5.2 Power comparisons for a simple linear regression model

We now investigate the power of our test in the simple case of linear trend plus an FD process. The model is

$$Y_t \quad = \quad \beta t + X_t, \qquad\qquad t = 0, \ldots, N-1,$$

where $\{X_t\}$ is a realization of an FD(d,1) process. The power of the test for trend was calculated via a Monte Carlo method based on 2048 random samples. Letting $q_{P,\alpha}$ denote the αth quantile of the test for trend statistic evaluated under H_0 by simulation (such as displayed Table 1), we estimated

$$\Pr(\text{reject } H_0 | \beta = \beta_1) \quad = \quad \Pr(P_c(d) > q_{P,0.95} | \beta = \beta_1),$$

for a range of slopes, β_1, until we reached a power of 1.

N	L	d	$p = 0.90$	se	$p = 0.95$	se	$p = 0.99$	se
256	4	0.0	1.089	0.001	1.101	0.001	1.126	0.001
		0.2	1.205	0.001	1.240	0.002	1.316	0.006
		0.4	1.612	0.006	1.741	0.008	2.055	0.020
		0.6	2.991	0.024	3.540	0.033	4.783	0.065
		0.8	7.353	0.092	9.533	0.151	14.887	0.329
512	4	0.0	1.048	0.001	1.054	0.001	1.067	0.001
		0.2	1.131	0.001	1.151	0.001	1.193	0.003
		0.4	1.472	0.004	1.571	0.007	1.799	0.015
		0.6	2.766	0.020	3.255	0.030	4.275	0.055
		0.8	7.159	0.082	9.061	0.139	14.02	0.291
1024	4	0.0	1.03	0.001	1.029	0.001	1.035	0.001
		0.2	1.09	0.001	1.098	0.001	1.128	0.002
		0.4	1.38	0.003	1.455	0.005	1.657	0.012
		0.6	2.63	0.018	3.073	0.028	4.075	0.061
		0.8	7.10	0.083	9.064	0.134	13.695	0.266
256	8	0.0	1.202	0.001	1.222	0.001	1.262	0.002
		0.2	1.414	0.002	1.455	0.002	1.556	0.005
		0.4	2.098	0.007	2.260	0.011	2.677	0.030
		0.6	4.646	0.033	5.462	0.052	7.375	0.142
		0.8	14.790	0.168	18.731	0.260	28.857	0.730
512	8	0.0	1.109	0.001	1.118	0.001	1.136	0.001
		0.2	1.257	0.001	1.281	0.002	1.336	0.004
		0.4	1.811	0.005	1.928	0.007	2.184	0.013
		0.6	4.130	0.029	4.783	0.045	6.424	0.111
		0.8	14.001	0.155	17.757	0.258	26.756	0.517
1024	8	0.0	1.059	0.001	1.063	0.001	1.072	0.001
		0.2	1.163	0.001	1.179	0.001	1.212	0.002
		0.4	1.622	0.004	1.712	0.006	1.929	0.012
		0.6	3.809	0.026	4.380	0.034	5.893	0.094
		0.8	13.632	0.152	17.444	0.242	26.407	0.683

Table 1: Table of the 90th, 95th and 99th quantiles of the centered test for trend distribution for various values of N, L and d. We picked the maximum value of J such that there was at least one nonboundary wavelet coefficient on level J. The evaluation of the distribution was based on 10000 simulations of an FD process using the Davies–Harte algorithm.

We compared the power of the DWT based test for trend with that of the standard linear regression test adjusted for long memory (see, e.g., Yajima (1988, 1991) and Deo and Hurvich (1998)). The OLS estimate of β (in the absence of an intercept) is given by

$$\widehat{\beta} \;=\; \frac{\sum_{t=0}^{N-1}(t-\bar{t})Y_t}{\sum_{t=0}^{N-1}(t-\bar{t})^2},$$

where $\bar{t} = N^{-1}\sum_{t=0}^{N-1} t$. By Theorem 1 of Deo and Hurvich (1998), for $d \in [0, 1.5)\backslash\{0.5\}$

$$N^{3/2-d}(\widehat{\beta}-\beta) \;\rightarrow_d\; \mathcal{N}(0, \sigma^2_{OLS,d}), \qquad \text{as } N \rightarrow \infty,$$

where

$$\sigma^2_{OLS,d} \;=\; \frac{72\sigma^2}{\pi}\int_{-\infty}^{\infty} \sin^2(x/2)\left[\frac{2}{x^2} - \frac{\cos(x/2)}{x\sin(x/2)}\right]^2 |x|^{-2d}\; dx.$$

The hypotheses of the test in this case are $H_0 : \beta = 0$ versus $H_1 : \beta > 0$. Letting q_α denote the αth quantile of the standard Gaussian distribution, the asymptotic power for this one-sided test is

$$Pr\left(\hat{\beta} > q_\alpha\, \sigma_{OLS,d}/N^{3/2-d} \middle| \beta = \beta_1\right)$$

where $\hat{\beta}$ has a $\mathcal{N}(\beta_1, \sigma^2_{OLS,d}/N^{3-2d})$ distribution. We could see no appreciable difference in the power of the test based on a asymptotic distribution of $\hat{\beta}$ compared with the Monte Carlo distribution of $\hat{\beta}$ when $N = 1024$.

Figure 8 displays the power curves for the following situations: (1) the OLS test adjusted for long memory (solid line); (2) the DWT based test for trend using a D(4) wavelet analyzing to $J = 8$ levels (dashed line); (3) the DWT based test for trend using a LA(8) wavelet analyzing to $J = 7$ levels (dotted line). In each case the time series was of length $N = 1024$. As d increases towards 0.6, the power of both wavelet tests improve relative to the OLS test. For $d = 0$ and $d = 0.2$ the Haar based test performs better than the D(4) test. For $d \geq 0.4$ the D(4) and LA(8) tests have similar power curves. The power is uniformly less for the wavelet tests compared with to the OLS test. This is not suprising since the wavelet based test for trend is nonparametric and has the potential to handle other types of trends.

In practice, estimating d will decrease the effectiveness of these tests. To counteract this we can alter our Monte Carlo testing scheme as follows. When we calculate the Monte Carlo distribution of $P(d)$, at each iteration we sample d from its limit distribution thus providing a test statistic which incorporates the variability in estimating the process parameters. A bootstrap sample estimate for d (Sabatini (1999)) could also be used.

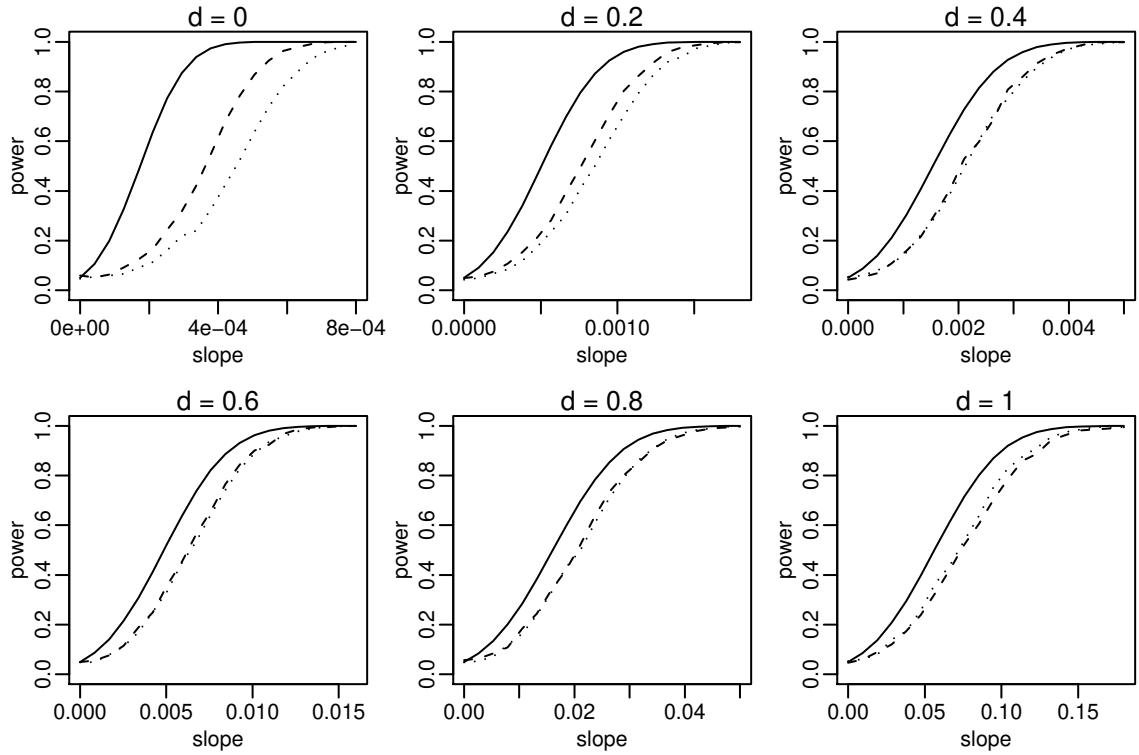

Figure 8: A comparison of the power curves for two wavelet based tests for trend (dashed line: D(4) wavelet filter; dotted line: LA(8) wavelet filter) and the least squares test (solid line) for various values of the slope, β, and difference parameter, d. The sample size used was 1024.

6 An example

Charles, Hunter, and Fairbanks (1997) present a 150 year record of the sea surface temperature (SST) from the Seychelles, in the Indian Ocean. Cores were collected from a coral colony growing at a depth of 7m, and the δ^{18}O oxygen isotope was measured in annual bands. The resulting monthly time series is shown in Figure 9. Following practice in the field we show the negative value of the isotope value. A decrease in the oxygen value corresponds to a increase in the SST temperature (roughly -0.24% corresponds to 1^{o}C). Charles, Hunter, and Fairbanks (1997) suggest that there is a significant decadal variability in this series which is commonly observed in areas with a monsoon climate. Thus, in this analysis we shall test for an interdecadal trend in the precense of what we will indicate is strong long range dependence. Our analysis will involve both a harmonic and wavelet based analysis of the time series. Concerning other prominent features in the data, Charles, Hunter, and Fairbanks (1997) attribute the peak in the data around 1878 to a monsoon failure in 1877. As

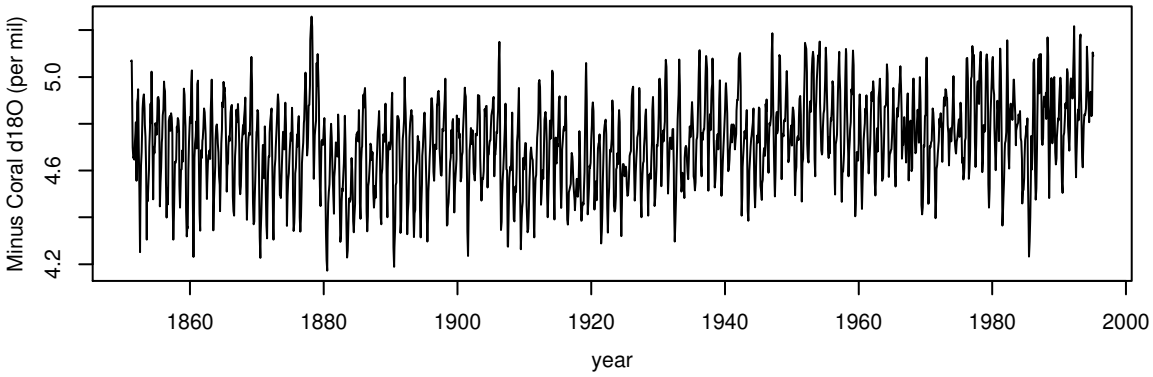

Figure 9: Time series of the negative monthly values of δ^{18}O oxygen isotope. The core sample was collected from a coral colony growing at a depth of 7m in the Seychelles, in the Indian Ocean.

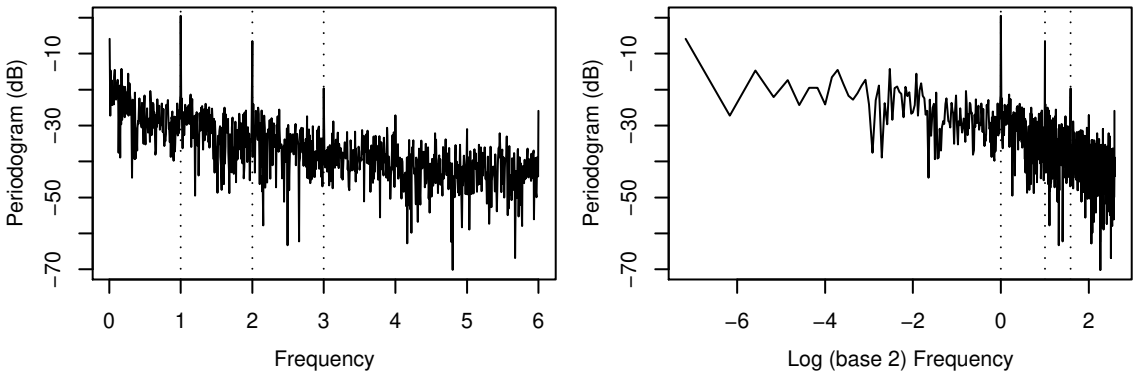

Figure 10: Left panel is a periodogram (in decibels) of the isotope series (after centering the series). Right panel shows the same periodogram on the $\log_2$ frequency scales. The dotted lines indicate the three significant periodicities in the data.

a consequence we would expect a local temperature increase in this period.

The left panel of Figure 10 shows a periodogram of the isotope series (in decibels). We can see significant peaks in the spectrum at $f = 1, 2$ and 3 corresponding to yearly, half-yearly and quarterly oscillations. The right panel of Figure 10 shows the periodogram of the data on the log-log scale. Using harmonic analysis, we remove the periodicities using harmonic regression. Remembering that this series has sample rate one month, i.e. $\Delta t = 1/12$, the log spectral density function is

$$\log S(f) \quad = \quad \log(2\sigma^2\Delta t) - 2d \log|2\sin(\pi f \Delta t)|, \qquad |f| < \frac{1}{2\Delta t}.$$

Since $\sin(\pi f) \approx \pi f$ for small f, an FD process is a good model for the data if the log spectrum

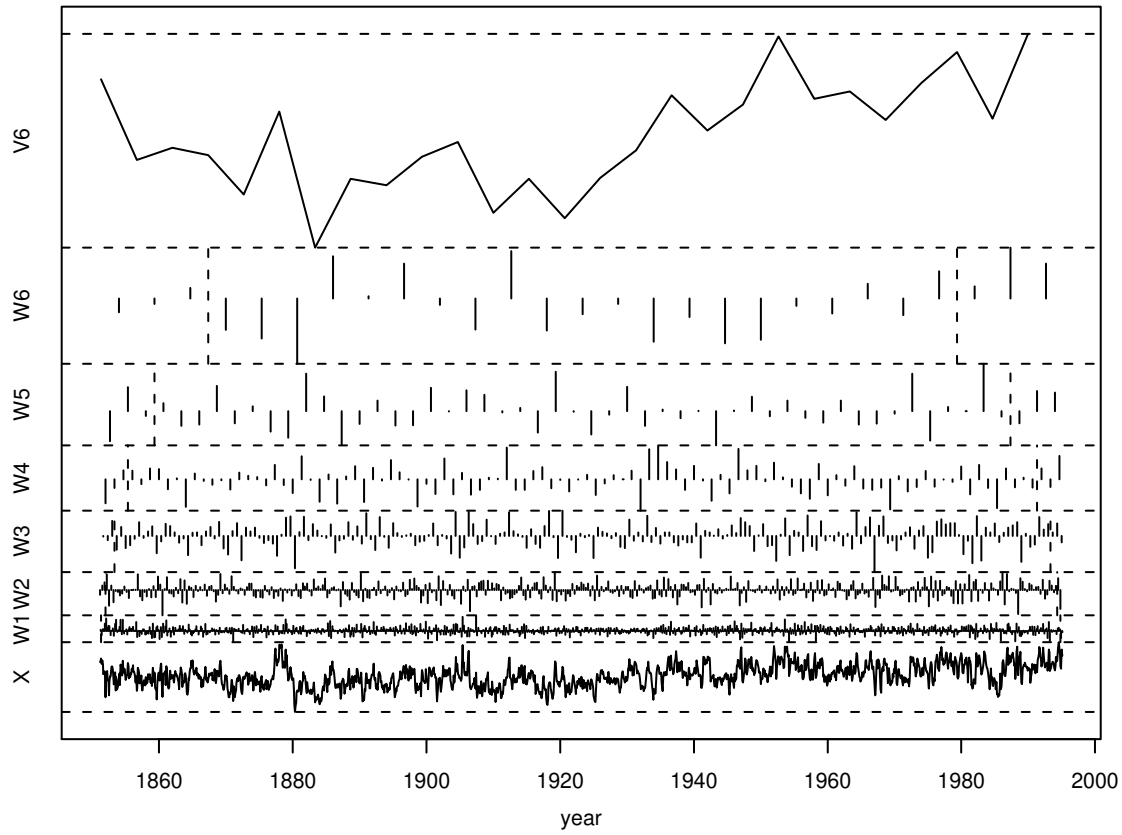

Figure 11: DWT of the data (after harmonic regression) using an LA(8) wavelet filter, analyzing to 6 wavelet scales (i.e., scaling coefficients correspond to averages over 64 months). Dashed vertical lines separate the boundary coefficients (outside) and nonboundary coefficients (inside). Note the monsoon failure in 1877.

versus log frequency is approximately a straight line, as in this case. By estimating the slope of the regression line we obtain an estimate of d. The estimate of $\hat{d} = 0.780$ based on a regression for $f \in [1/8, 5]$ indicates evidence of nonstationary long memory dependence.

We now study the series using the DWT. Figure 11 shows a DWT analysis of the data (after harmonic analysis) using a LA(8) wavelet filter. This filter allows for the possibility of a fourth order polynomial trend. We analyze to 6 scales so that the scaling coefficients correspond to averages over $64/12 = 5.33$ years. Thus the variability trend estimate approximately isolates the half decade and interdecadal variations. The vertical dotted lines on the plots denote the boundary between the nonboundary (inner) and boundary (outer) wavelet coefficients. Using the methods of

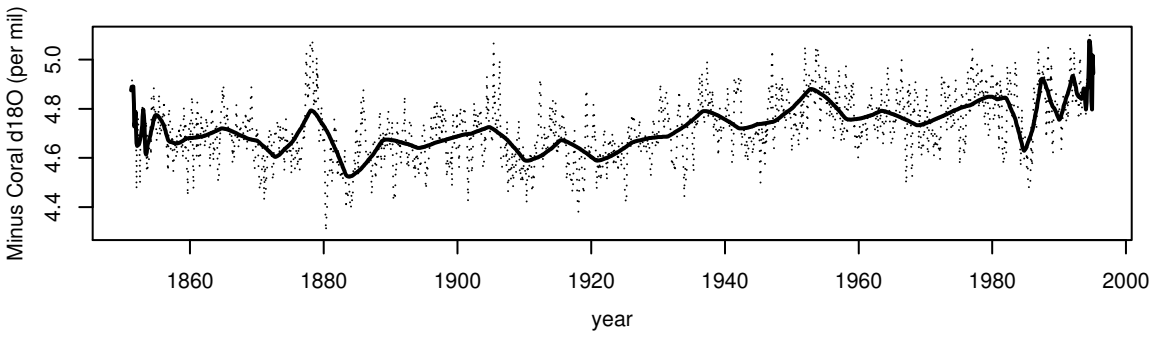

Figure 12: Solid smoothed line is a plot of the variable bandwidth estimate using an LA(8) wavelet filter analyzing to $J = 6$ wavelet levels. The dotted lines are a plot of the series after the 3 periodocities were removed using harmonic regression.

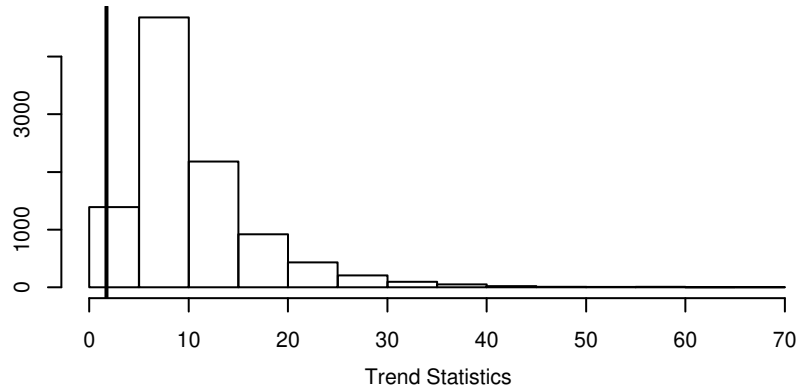

Figure 13: Distribution of the test for trend statistic. The solid line denotes the observed statistic.

Craigmile, Percival, and Guttorp (2000b) we obtain parameter estimates of $\hat{d} = 0.84$ with a 95% CI of [0.78,0.89]. The estimate of σ^2 is 0.005. This supports the idea of strong nonstationary long memory dependence.

We now investigate the half-decade and interdecadal variations, via the variable bandwidth trend estimate. Figure 12 shows the estimate as the solid line, along with the dotted lines which denote the series after harmonic regression. We can see that the monsoon failure affects the trend estimate locally around 1877. Overall, the trend estimate seems to adequately represent the large scale variations in the series. If we apply our test for trend (adapted to take into account the nonzero mean of the time series), we find that we cannot reject the null hypothesis of no trend. Figure 13 compares the value of the test statistic $P_c(d)$ for the time series (the solid vertical line) with the

distribution of the statistic under the null hypothesis. Since we would reject the null hypothesis when the observed $P_c(d)$ is in the upper tail of this distribution, there is no evidence for a trend. We can conclude that the large scale variations in the isotope series can be attributed to stochastic variations rather than to a deterministic trend. This conclusion is not surprising given the nonstationary value for d. Differencing the time series to make the error process stationary does not affect the conclusion of the test for trend.

Cole, Dunbar, McClanahan, and Muthiga (2000) analyze a 194 year record of the sea surface temperature (SST) based on a measurement of the δ^{18}O oxygen isotope from a coral sample in Malinda, Kenya. It would be of interest to compare the trend estimates for this series with the Seychelles trend estimates. We could also contrast these trend estimates to the El Niño Southern Oscillation 3.4 index. This will require modeling the correlations present between the three time series.

7 Discussion

In this paper we have considered DWT based methods of trend estimation and trend assessment. A clear advantage over traditional trend estimation methods for long memory processes is that it can be used simultaneously with DWT based methods of long memory estimation such as presented in Craigmile, Percival, and Guttorp (2000b, 2000a).

We have only considered a trend estimate based on a DWT of the data with Daubechies wavelet filters. Such wavelet filters are well suited to decomposing trend because of the inherent differencing and averaging operations. It would be possible to use other filters, but this would we compromise on the number of differencing operations. For example, the family of coiflets wavelet filters (Daubechies (1992), Percival and Walden (2000)) has $L/3$ differencing operations compared to the $L/2$ differencing operations of the Daubechies family. Also for $L > 2$, the Daubechies wavelet filters we use are not symmetric. This induces an asymmetry in the trend estimate which is not appealing in practice. One way to reduce artifacts attributable to asymmetric filters would be to average the estimate with the estimate obtained if we reversed the time series.

We can consider extensions to the methodology. All these results extend to autoregression frac-

tionally integrated moving average error processes (see section 2.5 of Beran (1994)). We induce more variability in the testing scheme as a result of estimating more parameters in the model. Even though in practice our scheme is designed to deal with low order polynomials, which are good approximations to many trends of practical interest, we conjecture the scheme can handle departures from this assumption using some appropriate asymptotic arguments. It is also possible to handle jumps in the time series. By noting that we can associate specific wavelet coefficients with time points we can remove the coefficients around the jump from the error process estimate $\widehat{\mathbf{X}}$ and instead add them to the trend component, $\widehat{\mathbf{T}}$. In estimating the parameters of the FD process, we also remove the wavelet coefficients around the jump points. To do this efficiently, we must know the location of the jump with some accuracy, and as a result make the trend estimate more erratic (with increased variability) about the jump point. See Wang (1998) for wavelet based methodology on change point detection.

We finish by noting that it is possible to obtain smoother estimates of trend either by increasing the number of wavelet scales we analyze (i.e., increase the value of J in the DWT) or by increasing the order of the wavelet filter (i.e., increase L). In the latter case we compromise on the accuracy of estimates of the parameters of the FD process because there will be fewer nonboundary wavelet coefficients (Craigmile, Percival, and Guttorp (2000b, 2000a)).

Acknowledgments

The authors gratefully acknowledge support for this research from an STTR grant from AFOSR (Insightful, Inc., and the University of Washington), an NSF grant (University of Washington) and an EPA grant (National Research Center for Statistics and the Environment). The authors also acknowledge comments from Prof. Noel Cressie which helped improve this paper.

References

Beran, J. (1994). *Statistics for Long Memory Processes*, Volume 61 of *Monographs on Statistics and Applied Probability*. New York: Chapman and Hall.

Beran, J. (1999). Estimating trends, long range dependence and non-stationarity. Technical report,

University of Heidelberg.

Brillinger, D. (1994). Some river wavelets. *Enivronmetrics 5*, 211–220.

Brillinger, D. (1996). Some uses of cumulants in wavelet analysis. *Nonparametric Statistics 6*, 93–114.

Charles, C. D., D. E. Hunter, and R. G. Fairbanks (1997, August). Interaction between the ENSO and the Asian monsoon in a coral record of tropical climate. *Science 277*, 925–928.

Cole, J. E., R. B. Dunbar, T. R. McClanahan, and N. A. Muthiga (2000, January). Tropical pacific forcing of decadal SST variability in the western Indian ocean over the past two centuries. *Science 287*, 617–619.

Craigmile, P. F. (2002). Simulating a class of stationary Gaussian processes using the Davies–Harte algorithm, with application to long memory processes. *to appear in Journal of Time Series Analysis*.

Craigmile, P. F., D. B. Percival, and P. Guttorp (2000a). The impact of wavelet coefficient correlations on fractionally differenced process estimation. In *European Congress of Mathematics: Barcelona, July 10–14, 2000, Volume II*, Progress in Mathematics (Volume 202), Basel, pp. 591–599. Birkhäuser Verlag.

Craigmile, P. F., D. B. Percival, and P. Guttorp (2000b). Wavelet-based parameter estimation for trend contaminated fractionally differenced processes. Technical Report 47, National Research Center for Statistics and the Environment, University of Washington.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Number 61 in CBMS-NSF Series in Applied Mathematics. Philadelphia: SIAM.

Davies, R. B. and D. S. Harte (1987). Tests for Hurst effect. *Biometrika 74*, 95–101.

Deo, R. S. and C. M. Hurvich (1998). Linear trend with fractionally integrated errors. *Journal of Time Series Analysis 19*, 379–397.

Draper, N. R. and H. Smith (1998). *Applied Regression Analysis (Third Edition)*. New York: John Wiley and Sons.

Gilbert, S. D. (1999). Testing for the onset of trend using wavelets. *Journal of Time Series Analysis 20*(5), 513–526.

Granger, C. W. J. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis 1*, 15–29.

Johnstone, I. M. (1999). Wavelet shrinkage for correlated data and inverse problems: Adaptivity results. *Statistica Sinica 9*, 51–83.

Johnstone, I. M. and B. W. Silverman (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society, Series B, Methodological 59*, 319–351.

Kendall, M. (1973). *Time Series*. London: Charles Griffin.

Künsch, H. (1986). Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability 23*, 1025–1030.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11*(7), 674–693.

Nicholls, D. F., C. R. Heathcote, and R. B. Cunningham (1986). The evaluation of long term trend I. *The Australian Journal of Statistics 28*, 294–313.

Percival, D. and A. Walden (1993). *Spectral Analysis for Physical Applications*. Cambridge: Cambridge University Press.

Percival, D. and A. Walden (2000). *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press.

Sabatini, A. M. (1999, December). Wavelet-based estimation of 1/f-type signal parameters: confidence intervals using the bootstrap. *IEEE Transactions on Signal Processing 47*(12), 3406–3409.

Sibbertsen, P. (2003). Log-periodogram estimation of the memory parameter of a long-memory process under trend. *Statistics & Probability Letters 61*, 261–268.

Smith, R. L. (1989a). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science 4*, 367–377.

Smith, R. L. (1989b). Reply to comments on "Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone". *Statistical Science 4*, 389–393.

Smith, R. L. (1993). Long-range dependence and global warming. In *Statistics for the Environment*, pp. 141–161. Wiley (New York).

Teverovsky, V. and M. Taqqu (1997). Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator. *Journal of Time Series Analysis 18*, 279–304.

Wang, Y. (1998). Change curve estimation via wavelets. *Journal of the American Statistical Association 93*, 163–172.

Wood, A. T. A. and G. Chan (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics 3*, 409–432.

Yajima, Y. (1988). On estimation of a regression model with long-memory stationary errors. *The Annals of Statistics 16*, 791–807.

Yajima, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *The Annals of Statistics 19*, 158–177.